BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 2: e26084
doi: 10.3897/biss.2.26084

OPEN ACCESS

Conference Abstract

Data Quality - Whose Responsibility is it?

Arthur D. Chapman ‡

‡ Australian Biodiversity Information Services, Ballan, Australia

Corresponding author: Arthur D. Chapman (biodiv_2@achapman.org)

Received: 23 Apr 2018 | Published: 13 Jun 2018

Citation: Chapman A (2018) Data Quality – Whose Responsibility is it? Biodiversity Information Science and Standards 2: e26084. https://doi.org/10.3897/biss.2.26084

Abstract

The quality of biodiversity data is an on-going issue. Early efforts to improve quality go back at least 4 decades, but it has never risen to the level of importance that it should have. For far too long the push to database more and more data regardless of its quality has taken priority. So I pose the question - what is the use of having lots of data if 1) we don't know what its quality is, and 2) if much of it is not fit for use?

When data-basing of herbarium and museum collections began in the 1970s many taxonomists saw the only use of the data as being for taxonomic purposes. But as more and more data has become digitally available, so too the uses to which the data can be put. It has also become increasingly important that the data we have in our herbaria and museums be put to more uses to justify on-going support and funding.

But whose responsibility is data quality? To answer that I take you to general data quality principles – i.e. that the difficulty and the cost of improving the quality of the data increases the further you move from its source. Responsibility for data quality rests with everyone.

- Collectors of the specimens
- Database designers and builders
- Data entry operators
- Data curators and managers
- Those responsible for exchanging/exporting the data
- Data aggregators

© Chapman A. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

- Data publishers
- Data users

We all have responsibilities.

So, what can we each do to play our part? We need to work together at all levels of the data chain. We need to develop systems whereby feedback on quality from wherever it comes can be documented and fed back. It is no use continually making corrections to the data down the line if those corrections never get back to the data curators and data custodians. It is also of little use if the information fed back goes nowhere and nothing is done with it.

The TDWG Data Quality Interest Group is working on setting up standards and tools to help make this possible. We have developed a Framework for Data Quality, we have developed a set of core tests for data quality, and assertions for feeding information back to custodians and forward to users and is beginning a process to deal with vocabularies of value for biodiversity data.

Keywords

biodiversity, data entry, aggregators, publishers, data users, vocabularies, data quality

Presenting author

Arthur D. Chapman